ELECTROSTATIC MODELING OF PROTEIN AGGREGATION

Ram Vanam

Submitted to the faculty of Indiana University
in partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Bioinformatics in the School of Informatics of,
Indiana University
December, 2004

Accepted by the Graduate Faculty, Indiana University in partial fulfillment of the requirements for the degree of Master of Science

_____

(Dr. Paul L. Dubin)

_____

(Dr. Douglas G. Perry)

_____

(Dr. Barry B. Muhoberac)

Table of contents

Acknowledgements

I appreciate this opportunity given to me in expressing my appreciation to all the people who helped me in the completion of the thesis.

I would like to convey my heartfelt gratitude to Dr. Dubin for his intellectual support and stimulating work environment. It was with Dr. Dubin's guidance that I learnt to think. I would also like to thank Dr. Perry and Dr. Muhoberac for their supervision.

I am grateful to the help offered by Emek Seyrek, Reddy Ganta, and Kati Giger.

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

LIST OF ABBREVIATIONS

Abbreviation

BLG…… Beta Lactoglobulin

BLG A……. Beta lactoglobulin A

BLG B…….. Beta lactoglobulin B

DLS……... Dynamic light scattering

IEP……Isoelectric point

ABSTRACT

Ram Vanam M.S., Indiana University, December 2004. Electrostatic modeling of protein aggregation. Research Advisor: Paul L. Dubin

Electrostatic modeling was done with Delphi of insight II to explain and predict protein aggregation, measured here for β-lactoglobulin and insulin using turbidimetry and stopped flow spectrophotometry. The initial rate of aggregation of β-Lactoglobulin was studied between pH 3.8 and 5.2 in 4.5mM NaCl; and for ionic strengths from 4.5 to 500mM NaCl at pH 5.0. The initial slope of the turbidity vs. time curve was used to define the initial rate of aggregation. The highest initial rate was observed near pH < pI i.e., 4.6 (< 5.2). The decrease in aggregation rate when the pH was increased from 4.8 to 5.0 was large compared to its decrease when the pH was reduced from 4.4 to 4.2; i.e., the dependence of initial rate on pH was highly asymmetric. The initial rate of aggregation at pH 5.0 increased linearly with the reciprocal of ionic strength in the range I = 0.5 to 0.0045M. Protein electrostatic potential distributions are used to understand the pH and ionic strength dependence of the initial rate of aggregation. Similar studies were done with insulin. In contrast to BLG, the highest initial aggregation rate for insulin was observed at pH = pI. Electrostatic computer modeling shows that these differences arise from the distinctly different surface charge distributions of insulin and BLG.

I. INTRODUCTION

Increased of biomolecular structures coupled with rapid developments in computational speed and visualization tools has made it possible to model and visualize complex electrostatic features of proteins. The resulting study of electrostatic protein surfaces yields insight into protein interactions. The most important aspect of protein surface potentials is in the identification and prediction of binding sites for ligands of varying size and complexity. The behavior of some less studied proteins can be inferred, which could facilitate the development of new therapeutic drugs and enhance our understanding of complex molecular pathways.

I.A. Structural Bioinformatics

Bioinformatics has been referred to as an interdisciplinary science that involves both conceptual and practical tools for the understanding, generation, processing and propagation of biological information[1]. The rapid increase in the number of 3D macromolecular structures available in databases such as the Protein Data Bank (PDB) has lead to the emergence of a sub discipline of bioinformatics: *structural bioinformatics.* Structural bioinformatics is a novel branch of biology which uses computational methods of analysis with the aim of modeling the 3D structures of proteins and macromolecular complexes [2].

I.B. Protein Structure and Function

In order to function, proteins must fold into their native, three-dimensional conformations. This folding is a consequence of the arrangement of linear chains of amino acids that differ due to the chemical properties of their R-groups (side chains). Interactions between contiguous amino acids leads to local folding of the amino acids chain (secondary structure) and then subsequent folding to give the overall shape of the protein (tertiary structure). The function of any specific protein is inseparable from its three-dimensional structure.

Proteins interact with other molecules – small molecules, other proteins, nucleic acids, and so forth – and these interactions are the basis of biological function. Proteins play a key role in cell division, cellular communication, metabolism, thus enforcing the activities encoded by genes. Protein interactions are central to the understanding of any cellular processes, and the corresponding studies flood the literature. Forces governing protein interactions include hydrogen bonding, hydrophobic forces, van der Waals forces, and electrostatic forces. Electrostatic interactions play a key role in governing the role of structure many important macromolecular processes such as molecular recognition, enzyme catalysis and the folding of nucleoprotein complex.

II. BACKGROUND

II.A. Protein Electrostatics

Proteins contain charged and polar groups, and electrostatic interactions control important aspects of their structure and function. These are long range, directionless (scalar), non-specific coulombic interactions that are important in understanding protein folding, stability, and function. These forces exert their influence over several nanometers in contrast to other short range electrostatic forces like salt bridges (interaction between two charged residues at very low dielectric constants) which are well oriented and short range. It is now quite clear that electrostatic effects play a major role in enzyme catalysis, electron transfer, proton transport, ion channels, ligand binding, macromolecular assembly, and signal transduction. They contribute to Protein – DNA interactions which are essential to genetic regulation of transcription, replication, translation and recombination. Complementarity was recognized by Nakamura and Wada, 1985[3] who illustrated its role in between ligand–protein and protein-protein surface interactions.

Electrostatics play an important role in catalytic reactions of proteins. Efficient enzyme catalysis is based on the decrease in the free energy of the transition state, and is mediated by a well-designed active site structure and electrostatic field[4-5]. For example, Asp 32 in subtilisin is believed to stabilize the protonated form of His64 in the transition state complex and to accelerate the catalytic reaction[6]. Similarly, unusual pK_a values of ionizable residues have often been found at the active sites of enzymes. In hen egg-white

lysozyme, an unusual ionization behavior has been observed for the catalytic carboxylates of Glu35 and Asp52[7].

Many models have been used to study protein electrostatics. These range from complex microscopic models which include atomistic details in the calculation of electrostatic energies (e.g., density function calculations, hybrid quantum mechanics-molecular mechanics, molecular dynamics or Monte Carlo simulations), to comparatively simple and reliable macroscopic models with no atomistic details (e.g., Poisson-Boltzmann, modified Tanford-Kirkwood model, generalized Born model)[8-19] and, bridging the gap between two, semi-macroscopic models (e.g., dipolar models)**.** The Poisson–Boltzmann treatment constitutes one of the most fundamental approaches to treat electrostatic effects in solution. Simple electrostatic models for globular proteins based on the Poisson-Boltzmann equation or its linear approximation were put forward quite early[19-22]. In subsequent years, enhancements of the nonlinear Poisson-Boltzmann equation with multiple dielectric constants, and new methods to decompose the energy terms to remove the dependency on the lattice grid (into which the proteins are mapped), have been reported[23]. These models, involving the Poisson-Boltzmann equation (linearized or not) have led to quite accurate computation of electrostatic potential at the solvent-accessible molecular surface, to computation of reaction rates between molecules in the solution, and to computation of the free energy of association of macromolecules and its salt dependence.

Electrostatic potential visualization helps in the identification of functional sites at the surface of a protein[24]. Electrostatic potentials can be visualized either by displaying isopotential surfaces or by color coding the molecular surface according to potential

values. Structural alignments based upon electrostatic potentials can provide important information about protein functions[25]. They may also identify residues that are functionally important[26].

II.B. Protein Aggregation and Electrostatics

II.B.1 Definition and Importance

Aggregates are operationally defined by three characteristics. First, they exhibit

Word	Phenomenon	Degree of polymerization	Induction	Cause
Association	Multimer formation	Unique	Natural	Electrostatic Short range
Association/ Aggregation	with out loss of native state or solubility but no precise organization	finite	Concentration, Non-physiological conditions	Hydrophobic Electrostatic Partial desolvation
Association/ Aggregation	loss of solubility; reversible native state retained salting out isoelectric precipitation	Infinite	Salt Co-solvents	Desolvation
Aggregation	loss of solubility and native state due to denaturing conditions (often irreversible)	Infinite	Denaturing conditions	Exposure of internal groups

Table 1. Terminology of association/aggregation used in literature and forces involved.

poor solubility in aqueous or detergent solvents. Second, they appear in aberrant subcellular or extracellular locations. Lastly, they exhibit secondary or tertiary structure that are not found in the normal native state. The terms "association" and "aggregation" are used interchangeably based on the types of protein assembly. "Association", but not "aggregation", may be used to describe multimerization wherein a protein stays soluble in the solution and is biologically functional. "Association" or "aggregation" may both be used when the protein assembly involves no precise organization but there is no loss of native state or solubility. "Aggregation", but not "association", is employed when there is a loss of both solubility and native state due to at least partial unfolding. The process of association/aggregation can be reversible or irreversible. Reversible aggregation occurs for example when the protein in its native state is subject to salting out or isoelectric precipitation. Irreversible aggregation, on the other hand, involves formation of misfolded and unassembled proteins typically due to exposure to extremes of pH, temperature etc.

In vitro manifestations of aggregation of native folded proteins are salting out (insolubility due to excess protein concentrations) and isoelectric precipitation (when a protein has zero net charge at its pI). Examples of *in vivo* manifestations of protein aggregation are inclusion bodies[27] (structures that contain extremely high concentrations of aggregated protein, formed in cells either to express heterologus/mutant proteins or upon over-expression of some endogenous protein), and Amyloids[28] (deposits of protein in the form of fibrils).

Protein aggregation is a major problem in disease states, technological processes, and research. Disease states included type II diabetes[29], Parkinson's disease[30], and Alzheimer's diseases[30] resulting from aberrant folding or processing events. However, protein aggregation need not be pathogenic. For example, in yeast, cytoplasmic inheritance of aggregated prion proteins underlies the propagation of stable epigenetic traits that are not associated with any known pathology[31]. Other consequences of protein aggregation arise in the biotechnological development of drugs[32], the production of recombinant proteins[33], and in studies of protein folding and stability. The enormous impact of disease states and the growing number of protein drugs have given rise to intense research of both equilibrium and dynamic aspects of protein aggregation, with the ultimate objective of learning to manipulate the protein chemistry and system conditions in ways that will preclude or slow down disease process.

II.B.2 Protein aggregation and electrostatics

Electrostatics is one of a number of factors effecting aggregation. In particular, it is believed that screening of electrostatic repulsions is a prerequisite for aggregation because it was assumed that electrostatic interactions oppose protein-protein association processes. However, growing evidence suggests that charged groups can contribute favorably to the stabilization of protein complexes[32]. The rate of association is governed by diffusion and can be increased by favorable coulombic electrostatic forces[34-35]. Conversely, the rate of dissociation is dictated by the strength of short range interactions

between the proteins (van der waals interactions, hydrogen bonds, hydrophobic interactions and salt bridges)[36].

It is still not clear to what extent electrostatic interactions can contribute to the stabilization of protein-protein complexes. However, it is known that electrostatic fields around proteins can increase the rates of protein-protein association[37-39] and that electrostatic interactions have an effect on the specificity of protein-protein association. The role of electrostatics in protein-protein association can be manifested in binding strengths, specificity, or rates of association. For they account for an increase in the rates of protein association between TEM1 beta-lactamase and its protein inhibitor BLIP by 250-fold[36]. Upon incorporating charged residues in the vicinity of the BLIP binding surface. Electrostatics has been implicated in multimerization of human arylsulfatase A (ASA, EC 3.1.6.8) an enzyme that occurs in the lysosome of the human cell, where acidic pH values prevail[40]. ASA forms octamers in an acidic environment (pH $\leq$ 6) and dissociates to dimers at higher pH values, and the pH-regulated association between dimers of ASA has been explained on the basis of electrostatic interactions. Encounter complex formation between barnase and barstar is increased by an interactive long range force strong enough to influence the translational and rotational Brownian motion of the proteins before the complex is formed[41]. It has been shown by Ferhst and co-workers[42] that barstar looses its stability to rapidly associate with barnase in forming a cluster of negatively charged residues facing barnase. It was proposed that long-range electrostatic force could promote the formation of encounter complexes, which accelerate the rate of complex formation and lifetime of complexes[40,43].

II.B.3. Aggregation behavior of proteins under study

II.B.3.1.Beta Lactoglobulin

The structure of Bovine β-lactoglobulin (BLG), the most abundant protein in cow's milk whey, is well understood. Its physiological function is still a mystery[44], but it seems to be involved in transport of small, sparingly soluble molecules such as retinol or fatty acids, although species distribution and variation in binding profile do not all fit such a role[45]. The native state is a dimer of identical units, each unit containing 162 amino acid residues (18kD). Each monomer in it contains two disulfide bonds (Cys66-Cys160 and Cys106-Cys119) and a free thiol (Cys121). It is predominantly a β-sheet protein consisting of nine β-strands (A – I) and one major α- helix at the C-terminal. Some 10 variants of the BLG have been identified, but by far the most common are the A and B forms that differ in only two places: Asp64Gly and Vall118Ala[44].

Multimerization of BLG is electrostatic and associates readily at low ionic strength at pH close to its pI[46]. At room temperature, BLG exists as a monomer, dimer or can self-associate to an octamer[47-49]. The monomeric form predominates below pH 3[49-51] and above pH 9[53], but coexists in equilibrium with dimers in the pH ranges 2.0-3.7 and 5.2-9.0, whereas dimers and higher order aggregates were reported between pH 3.7 and 5.2[54-56]. These higher-order aggregates at pH 4.7 were reported to be octamers[50,54,57]. Ionic strength seems to have some influence on the aggregation of BLG, aggregation

decreasing with increase in ionic strength[53, 58-59]. The pH and ionic strength regulation of BLG association/dissociation suggests that it may be governed by electrostatic forces.

Electrostatics interactions play a significant role in BLG aggregation. In BLG as for other proteins, it is important to understand the forces that drive association and consequently to be able to regulate these forces by external parameters. A number of forces have been proposed to significantly influence the strength, specificity, and rate of association. Piazza and Lacopini[60] showed that a possible result of attractive interactions in BLG A solutions is the spontaneous formation of transient clusters but did not specify the nature of the interactions. Recent investigations showed that salt bridges stabilize the BLG dimer by decreasing the electrostatic repulsion between protomers at pH 3.0[49]. However, theoretical investigations showed that electrostatic interactions can have a stabilizing effect on protein protein binding[61].

II.B.3.2. Insulin

Insulin is one of the most intensely studied biomolecules in biochemistry. Its chemical sequence determined in 1955 was the first established for any protein[62-63] and its three dimensional structure of the molecule was determined by X-ray crystallography[64] in 1969. Insulin performs the role of increasing the uptake of glucose in the body cells upon detection of elevated sugar in the blood stream by pancreas. An inability to produce sufficient levels of insulin can result in diabetes mellitus. In 1967 it was discovered that insulin is synthesized in the B-cells of the Islets of Langerhans,

where a single chain of preproinsulin is converted to proinsulin, which is subsequently converted enzymatically to insulin in the storage vesicle[65].

Insulin is one of the most frequently studied self associating proteins, exhibiting a complex association pattern consisting of monomer, dimer, tetramer, hexamer, and higher-order species[66-74]. The 5800Da monomer consists of an A chain with 21 residues (which are denoted by A1-A21), and a B chain, 9 residues longer (B1-B30), held together by two inter-chain disulphide bonds (between A7 and B7, as well as A20 and B19), while an additional disulphide connects A6 and A11 within the A chain. This "monomer" multimerizes readily to form dimers, which are stable in aqueous solution in pH 2 and 8. In the presence of zinc and calcium ions, three such dimers assemble to form a hexamer of insulin monomers[66], which is stable between pH 5 and 8. This pattern of assembly is utilized in the biosynthesis, processing and storage of the hormone. Upon extracellular release, the hexamers dissociate into dimers and eventually into monomers which are the biologically active species. It has been found that the strongest subunit interactions are between dimers, and that if these interactions could be disrupted, then the association to hexamers could be stopped. Numerous models have been proposed to explain insulin self-association[75-76]; however, the molecular mechanism of insulin aggregation remains substantially unknown.

III. PROPOSAL

III.A. Knowledge gap

Protein aggregation is a complex phenomenon that can occur in vitro and in vivo, resulting in the loss of the protein's biological activity. Numerous studies have been carried on aggregation of beta lactoglobulin and insulin; but no specific studies focusing on the mechanism of aggregation have been proposed. More over a clear understanding of the role of electrostatics in the aggregation process is lacking.

III.B. Objectives

The objectives of this work were two fold.

1. The first was to investigate pH and ionic strength dependence of BLG aggregation, and to use computer modeling to understand the role of electrostatics in protein this aggregation.
2. The second was to use electrostatic computer modeling for predicting aggregation behavior of insulin.

III.C. Intended research project

1. pH dependence of initial rate of aggregation of BLG studied at a constant ionic strength of 4.5mM by colorimetry.
2. Ionic strength dependence of initial rate of aggregation of BLG studied at pH 5.0 by colorimetry.

3. pH dependence of aggregation rate for different genetic variants studied at constant ionic strength of 4.5mM by stopped-flow spectrophotometry.
4. Electrostatic modeling of BLG dimer to explain the pH and ionic strength dependence of initial rate of aggregation.
5. To predict aggregation behavior of insulin using electrostatic modeling.

IV. ORGANIZATION OF THESIS

The current work is a continuation of previous aggregation studies of BLG. An in depth understanding of the aggregation mechanism was achieved through experimental observations coupled with electrostatic modeling. In this thesis, further research on the aggregation of BLG including role of genetic variants was conducted using stopped-flow spectrophotometry, rather than conventional colorimetry. With a thorough understanding of the aggregation of BLG, the aggregation studies were extended to insulin. Insulin aggregation was studied using electrostatic modeling as a predictive base. BLG was written as Results followed by Discussion, in order to discuss colorimetry, QELS, modeling and theory together. For insulin we focus on difference from BLG which appears with simple pH and ionic strength studies from colorimetry.

V. **METHODS**

V.A. Materials

Most studies were carried with β-Lactoglobulin A&B (no. L-2506-5G, Lot 032K7035) from Sigma-Aldrich (St Louis, MO). Lot to lot variability was explored using lots 101K7031, 20K7023. To define the role of genetic variants, we used BLG A (no. L7880-25MG, Lot 031K7052) and BLG B (no. L8005-25MG, Lot 011K7032). Human insulin, LY41001(Lot KG5-WEF-179) was obtained from Eli Lilly and Company. pH was adjusted with 0.1N hydrochloric acid and/or 0.1 N sodium hydroxide solutions. All solutions were made from Milli-Q water (Millipore, Milliford, MA).

V.B Methods

V.B.1 Sample preparation

1 g/L BLG solutions were prepared in NaCl solution at appropriate ionic strength by stirring for at least 15 min and filtering with 0.22 μm filters (Sartorius AG, Germany) prior to use. Solutions were then adjusted to pH 9.0 to ensure formation of stable monomer. Two methods were employed to bring this solution to the desired initial pH.

V.B.2 Mixing technique

BLG at pH 9.0 was rapidly mixed with a predetermined solution of HCl in NaCl to bring the pH from 9.0 to the target pH. Similar sample preparation and mixing technique were also used in experiments carried out with insulin. The amount of acid added was negligible and its effect on ionic strength can be neglected. The target pH was attained within about 5 seconds.

V.B.3 Stopped-flow spectrophotometer

The stopped flow experiments were performed at room temperature (23°C) of BLG with a Hi-Tech Scientific SF-61SX2 kinetasyst stopped-flow spectrophotometer, equipped with a 75 W Xenon lamp. All measurements were made in absorbance mode at 420nm. A solution of BLG pH 2 g/L at pH 9.0 was mixed with an equal amount of 4.5mM NaCl solution containing the calculated amount of acid to bring the pH to a target value. The data was converted to 100 - %T before analysis.

V.B.4. Turbidimetric titration

Turbidity measurements of BLG or insulin reported as 100-%T, were performed at 420 nm using a Brinkman PC 800 probe colorimeter equipped with a 1cm path length probe, calibrated to 100% transmittance with Milli-Q water. The pH was measured with an Orion 811 pH meter equipped with a Beckman refillable combination pH electrode

and calibrated with pH 7 and pH 4 buffers. % T was monitored with time after the target pH was attained using the sample preparation technique described above.

V.B.5. Dynamic light scattering (DLS)

A DLS system from Brookhaven Instruments(Holtsville, NY) equipped with a BI-9000AT digital autocorrelator and a 488 nm, 100 mW argon-ion laser was used to measure the particle size of BLG at 90° scattering angle for pH values 5.2 and 5.4. Measurements at pH 4.0 and 4.2. Measurements every 30sec were carried out with a Malvern zetasizer (Malvern Instruments). The mean apparent translational diffusion coefficients were obtained by fitting the autocorrelation function using NNLS (Brookhaven) or a " "(Malvern). The hydrodynamic radius (R_h) of the particles was determined from Stokes-Einstein equation.

$$R^h = K_b T / 6\pi\eta D_T \quad (1)$$

where K_b is Boltzmann's constant, T is the temperature , and η is the solvent viscosity.

V.B.6. Initial rate determination

Initial rate of aggregation for BLG or insulin was determined from the initial slopes of turbidity vs. time curves. The first ninety seconds data was used for determination of rates for both BLG and insulin. The rates were determined using Microsoft excel. This is illustrated in figure 2 and described in detail.

V.B.7. Computational methods

Computer modeling allows visualization of the electrostatic potential around the protein as a function of pH and ionic strength. In Delphi V98.0 (Molecular Simulations Inc), the electrostatic potential around the protein is calculated by nonlinear solution of the Poisson–Boltzmann equation. The protein crystal structures with Protein Data Bank identification 1BEB.pdb were taken from RCSB Protein Data Bank (http://www.rscb.org). The deposited structure 1BEB.pdb has A variant Val at position 118, and B variant Gly at 64. This corresponds neither to BLG A (Asp64, Val118) nor to BLG B (Gly64, Val118). In order to rectify this incorrect amino acid sequence, the charge file used for the electrostatic calculations was modified by replacing the Gly64 with Asp64 to mimic a BLG A dimer. The BLG B dimer was not considered because the association of BLG A dimer appears to dominate aggregation effects. Modeling studies for insulin were carried out with dimer (1ZNI.pdb). The charges of the amino acids in the proteins were determined using the spherical-smeared –charged model put forward by Tanford[77] utilizing the protein titration curve of BLG[78] and insulin[79].

VI. **RESULTS AND DISCUSSION**

VI.A. β LACTOGLOBULIN

VI.A.1 RESULTS

VI.A.1.1.*Effect of pH*

Figure 1 shows the Turbidimetric response of 1g/L BLG in 4.5mM NaCl at various initial pH values obtained by colorimetry. The pH drift during the course of the experiment was ≤ 0.01 pH units. Time zero corresponds to the time at which the target pH was attained; at this point the transmittance was equal to 100 ± 0.2%. In the pH range 4.3 – 5.0, the abrupt initial increase in turbidity is followed by a stable state, but a continual increase was observed for the more extreme pH's. The initial aggregation rates expressed as $(d\tau/dt)_0$, were obtained as the initial slopes turbidity vs. time as shown in figure 2, an expansion of the data in figure 1. As an example, the line drawn for the pH = 4.59 data

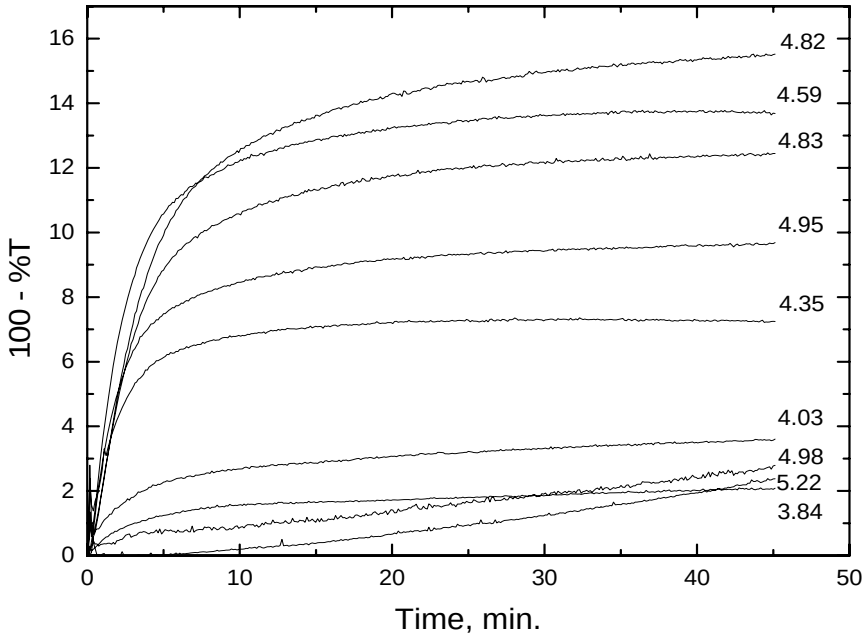


Figure 1. Turbidity vs. time for BLG 1g/L, I = 4.5 mM at pH values shown. Data obtained by probe colorimeter.

corresponds to the linear best fit by excel for the time interval of 0 – 90 seconds, including the origin, and automatically rejecting obvious outliers. This procedure was followed in call cases to get $(d\tau/dt)_0$. Extending the data set to 0 – 120 seconds never effected $(d\tau/dt)_0$. It should be noted that this protocol differs from the earlier work[80]. In earlier work the first two minutes turbidimetric data has been taken for calculation of initial rates manually. The slopes were determined by taking the best linear fit. However, the initial outliers were rejected and time zero was not considered in determination of the rates.

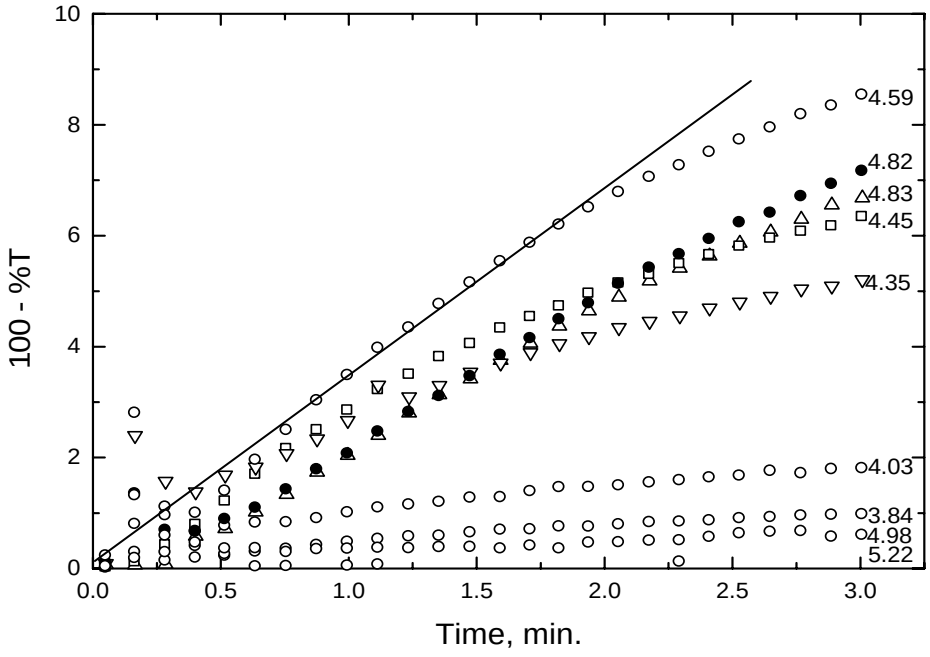


Figure 2. Expanded scale for figure 1 to show method of obtaining initial slope.

The pH dependence of BLG aggregation was also monitored using stopped-flow spectrophotometry. The data were recorded as absorbance and converted to 100 - %T. Similar to the treatment colorimetric data, the initial slopes of 100-%T vs. time were used to determine the pH dependence of initial rate of aggregation. Figure 3 shows the pH dependence of the initial aggregation rate obtained by both stopped-flow and colorimetry data. In both data sets, the highest rate of aggregation was observed at pH 4.6. In contrast to the colorimetry data, aggregation rates were not close to zero at pH's 5.0 – 5.2. In the absence of duplicate runs for any given pH, the precision is unknown; some data suggest high precision of ± 0.1, but other data show deviations of ± 1. Between pH 4.1 – 4.45 and 4.7 – 4.85 agreement is good between the two methods. However data points at pH's 4.6

and 5.0 determined by stopped-flow seem to be outliers. Neglecting these outliers, data obtained by both colorimetry and stopped-flow photometry are consistent with a single curve of pH dependence of initial rate of aggregation.

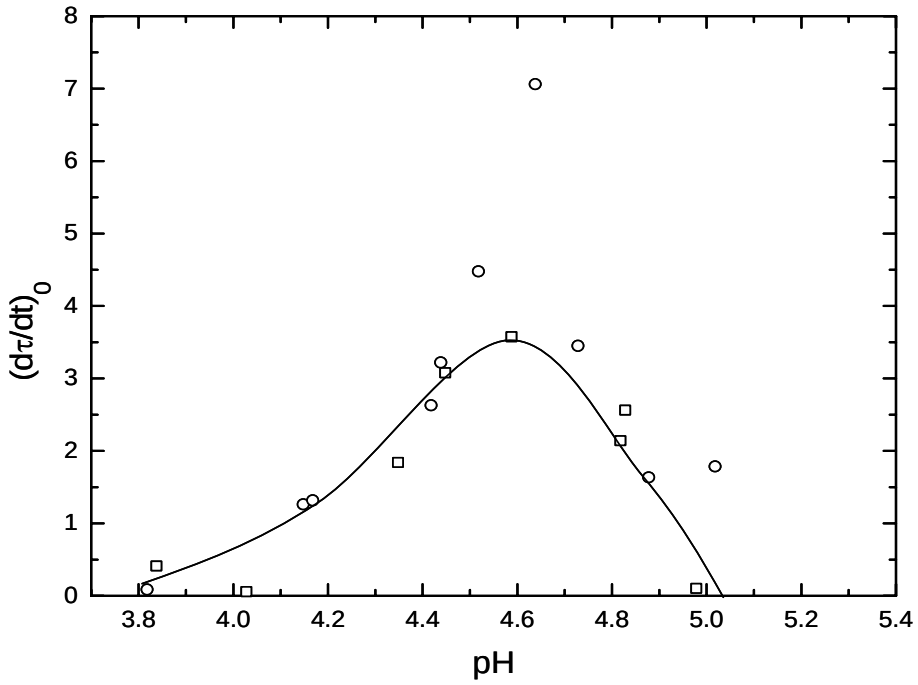


Figure 3. Initial Rate vs. pH for 1g/L BLG at 4.5 mM ionic strength for data obtained from turbidimetry (□) and stopped-flow photometer(O).

VI.A.1.2.*Effect of ionic strength*

Figure 4 shows the turbidimetric response of 1g/L BLG at pH 5.0 for various ionic strengths ranging from 0.0045 to 0.5M. Minor changes in turbidity with time were observed at I>0.02M. At I<0.1M a sharp increase in turbidity was observed in the first few minutes. Ionic strength dependence of initial rate of aggregation was determined from the slopes of initial few minutes data, similar to the pH dependence.

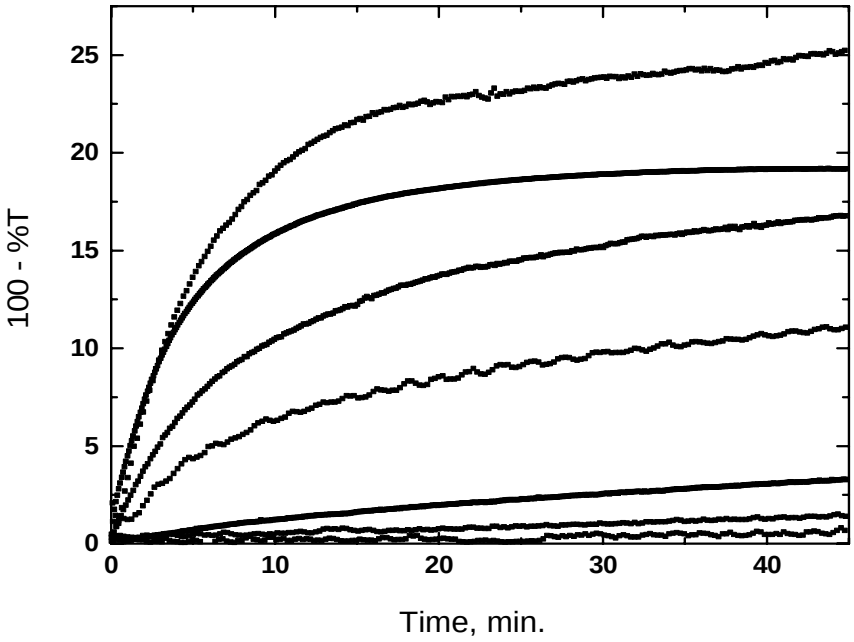


Figure 4. Turbidity vs. time for 1g/L BLG, pH 5 at different ionic strengths (top to bottom I = 0.0045M, 0.006M, 0.008M, 0.02M, 0.1M, 0.3M, 0.5 M). Data obtained by probe colorimeter.

VI.A.1.3.*Sample Variability*

When the ionic strength dependence of initial rates of aggregation was examined using different lots of BLG by colorimetry, the data displayed drastic changes in the absolute rates. To further investigate this lot-to-lot variability, the pH dependence of the initial rate at I =4.5mM was obtained. As shown in Figure 5, all three lots showed the same asymmetry of the pH effect. However, the pH for maximum aggregation ranged from 4.6 to 5.2.

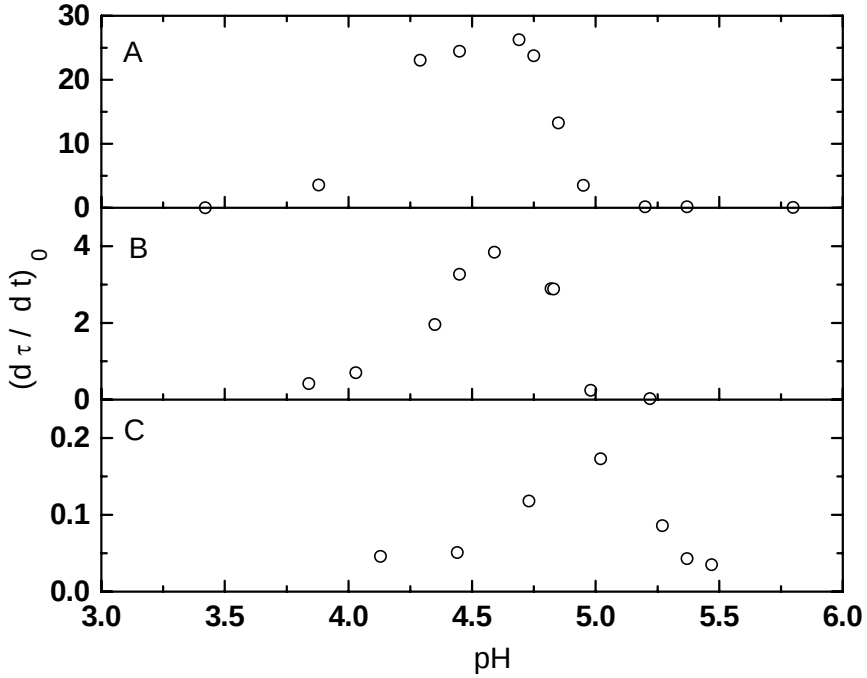


Figure 5. Initial rate vs. pH for 4.5mM, 1g/L BLG from different (SIGMA) lots. Lot 20K7023(A); Lot 032K7035 (B); Lot 101K7031 (C). Data obtained by probe colorimeter.

VI.A.1.4.*Genetic variance*

Figure 6 shows turbidimetric response of BLG variants obtained by stopped-flow spectrophotometry. In experiments I and IV 0.5 g/L of BLG A or BLG B solutions at pH 9.0 were rapidly brought to a target pH of 4.8. The resultant solution at pH 4.8 would have dimers of BLG A or BLG B as the initial reactants. Results in figure 6 show rapid aggregation for BLG A, (I) and negligible aggregation for BLG B (IV). In II BLG A and B were mixed at pH 9.0 to yield a solution of 1 g/L in total protein concentration and then brought rapidly to pH 4.8. The resultant solution would have dimers of AA, BB and AB as the initial reactants at pH 4.8. Significant aggregation was noticed in (II), however was

lower compared to (I). In III BLG A and B were first dissolved at pH 6.0 to yield a solution of 1g/L and then mixed at that pH, and brought to 4.8, resulting in dimers of BLG A and B as the initial reactants. Results in figure 6 show negligible aggregation for III.

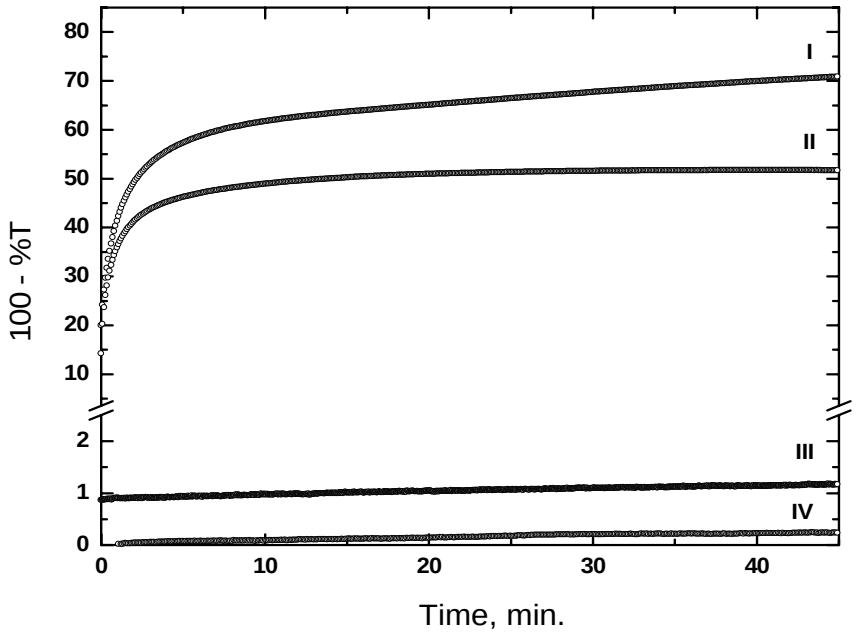


Figure 6. 100 - %T vs. Time for BLG genetic variants A, B and AB at I = 4.5 mM. (I) 0.5 g/L BLG A at pH 4.8; (II)1g/L final concentration BLG A+B mixture at pH 4.8 brought down from an initial of pH 9; (III) 1g/L final concentration BLG A+B mixture at pH 4.8 brought down from an initial of pH 6 ; (IV). 0.5g/L BLG B at pH 4.8. Data obtained by stopped-flow spectrophotometry.

VI.A.1.5.*Time dependence of particle size*

Figure 7 shows the time dependence of particle size for 1g/L BLG in 4.5mM NaCl at pH 4.2, monitored by QELS. QELS results reveal a distinct 6 nm mode at all

times, which shows the continuous presence of dimers. Two distinct higher order modes of 600 – 700 and 200 – 400 nm were also identified at pH 4.2.

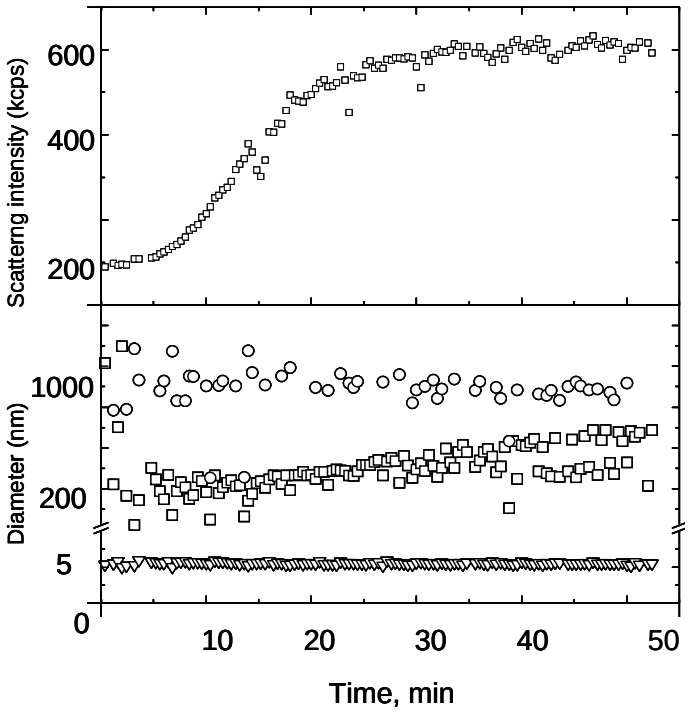


Figure 7. Time dependence of QELS for 1 g/L BLG; I = 0.0045M; pH 4.2. Time dependence of total count rate (upper); apparent diameters (middle); and absolute scattering intensity corresponding to each mode (lower). Symbols in lower curves identify fast (▽), slow (λ) and intermediate (▽) modes, corresponding to the diameters 6 nm, 200-400 nm, and >600 nm, respectively[80].

VI.B.1.DISCUSSION

VI.B.1.1..*Effect of pH*

The observation of maximum initial rate of aggregation at pH 4.6, with extreme pH sensitivity between 4.6 and 4.9 can be attributed to asymmetry of charge distribution. The asymmetry observed could possibly be due to the large changes in the aggregation rate resulting from smaller changes in the pH. The asymmetry of charge and potential distributions lead to attractive forces between positive and negative domains of associating proteins. These local distributions are strongly influenced by the charge state of particular amino acids. These effects are best visualized by electrostatic computer modeling, which has been used to understand similar inter-macromolecular phenomena related to protein charge heterogeneity[36,82.]

Electrostatic modeling was done was done with BLG structure (1BEB.pdb) in the previous studies[80] which has incorrect amino acid sequence of A variant Val at position 118; and B variant Gly at 64 which corresponds to neither BLG A (Asp64, Val118) nor BLG B (Gly64, Ala118). In order to rectify the incorrect amino acid sequence, the charges on the structure were modified by replacing the charge of Gly 64 with that of Asp 64, to mimic a BLG A dimer. Modeling the deposited structure would give the electrostatic models for BLG B dimer because it has the amino acid residues with charges corresponding to the BLG B dimer.

Figure 8 shows potential distributions for the AA dimer to understand the pH dependence of initial rate of aggregation observed in figure 3. The negligibly small initial rates below pH 4.0 and above pH 5.0, as seen in Figure 3, are due to the net positive and

negative protein charge respectively, preventing association. At pH 5.0, close to pI, BLG shows approximately equal positive and negative lobes, as seen in Figure 8 D. Many proteins exhibit diminished solubility near the isoelectric point; this phenomenon, often referred to as "isoelectric precipitation"[82], is frequently accounted for in terms of minimal protein net charge corresponding to minimization of aggregation-inhibiting repulsion. However, it is evident here from modeling that (a) aggregation is promoted by electrostatic attractive forces, and (b) the net protein charge is not highly relevant to these forces. A net protein charge of zero at pH 5.2 and positive and negative lobes of similar magnitude do not constitute the condition of maximum aggregation, but instead occurs at pH 4.6 where electrostatic potential contours are highly asymmetric.

While we can visualize how positive and negative domains of similar size could support one-dimensional association, the ability of the large positive lobe in Figure 8 B to accommodate the small negative domains of several other proteins suggests the multidimensional propagation that is required to account for large aggregates. The former leads to linear chains of finite length, and the latter to branching to infinite molecular weight. This description is analogous to the difference between condensation polymerization of A-B type vs. A-$(B)_2$ monomers[83].

The possibility of ion-binding, particularly as the binding of chloride ion would be expected to contribute to asymmetric potential contours. In order to establish whether our neglect of chloride ion binding could introduce a significant error in our modeling we repeated calculation after removing two positive charges, corresponding to the maximum number of chloride ions bound to BLG. No significant effect on the computed positive

domain was observed. This enhances our confidence that the large positive domain at the pH of maximum aggregation is not susceptible to diminution by ion binding.

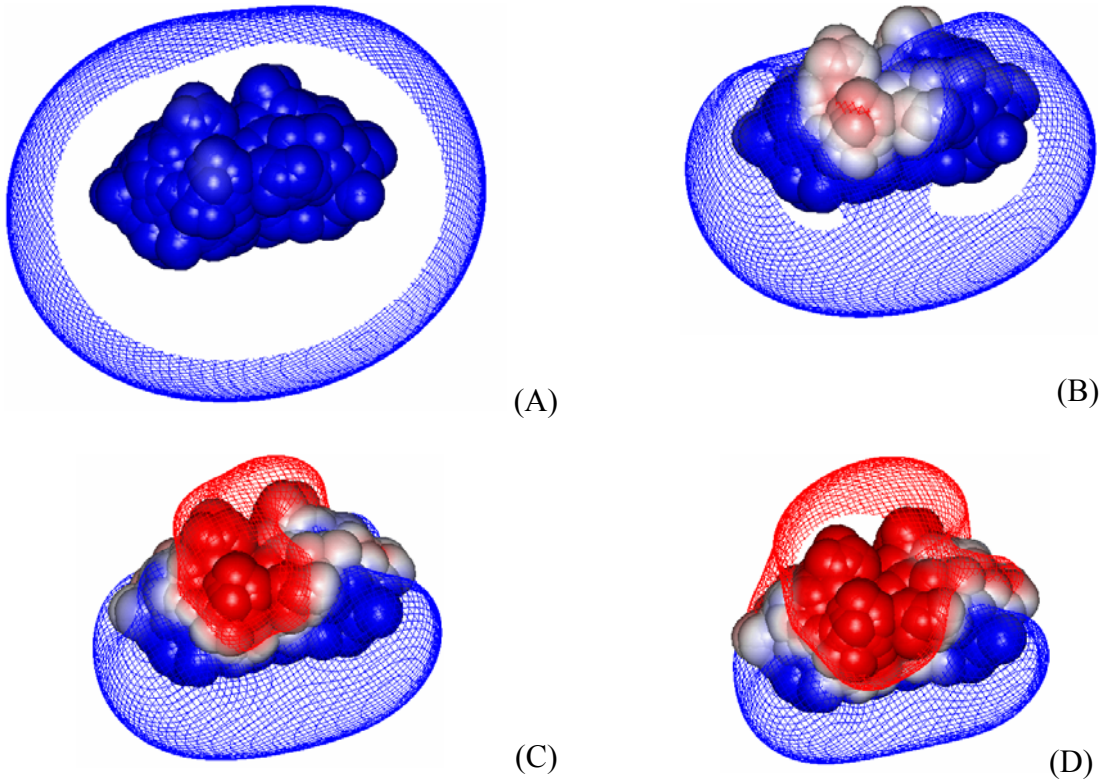


Figure 8. 0.5(blue) and -0.5(red) kT/e electrostatic potential contours around BLG AA dimer at ionic strength 0.0045M at pH (A) 4.03, (B) 4.59, (C) 4.98, (D) 5.22

VI.B.1.2.*Effect of ionic strength*

Figure 9 shows the ionic strength dependence of initial aggregation rate for 1g/L BLG at pH 5.0, obtained by colorimetry. A significant increase in initial aggregation rate was observed with decrease in *I* from 0.1 to 0.0045M. As shown in the insert of Figure 5,

the data are fit well by $(d\tau/dt)_0 = 0.017\ I^{-1}$ .The significance of this result will be discussed later.

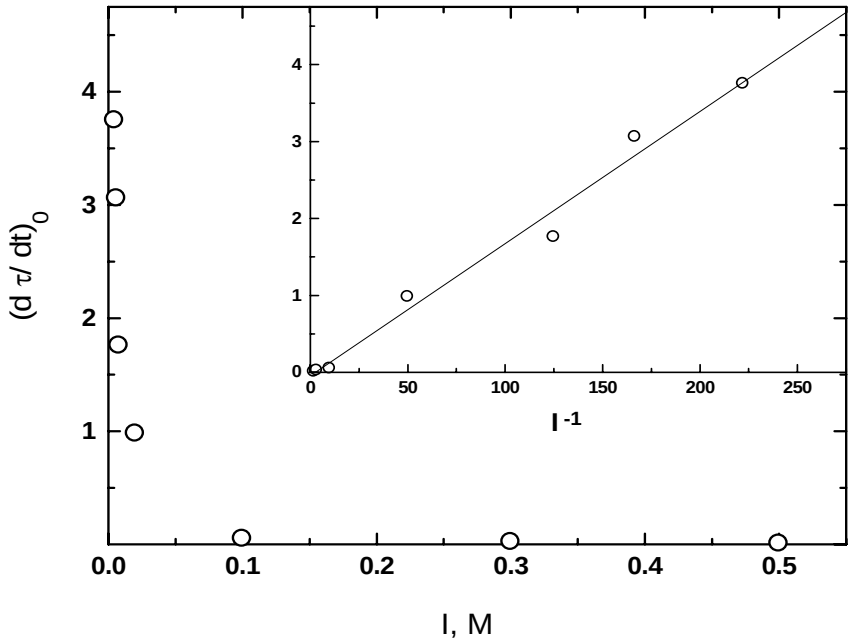
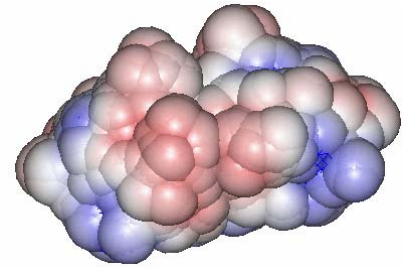
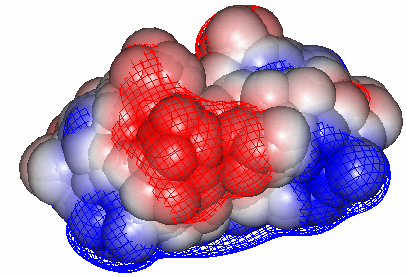


Figure 9. Initial rate vs. *I* for 1g/L BLG at pH 5.0. Insert: Initial rate vs. I^{-1}

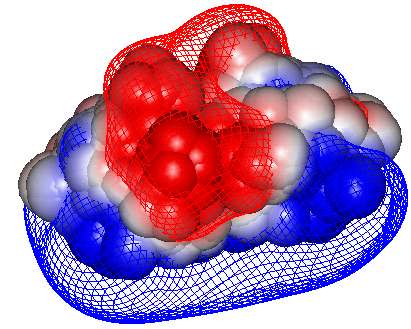
Figure 10 addresses the ionic strength dependence of aggregation by electrostatic modeling for BLG A dimer. The small variation in the electrostatic domains for ionic strengths 0.5 and 0.1 M as seen in Figure 10 account for the negligible differences in the initial rates at these ionic strengths. As *I* decreases below 0.1M, the most dramatic effect is the expansion of the positive domain, and the appearance of strong asymmetry between positive and negative regions with regard to both shape and magnitude. Figure 10D strengthens the hypothesis that BLG aggregates most when it has a distinct negative domain and a dominant positive domain.



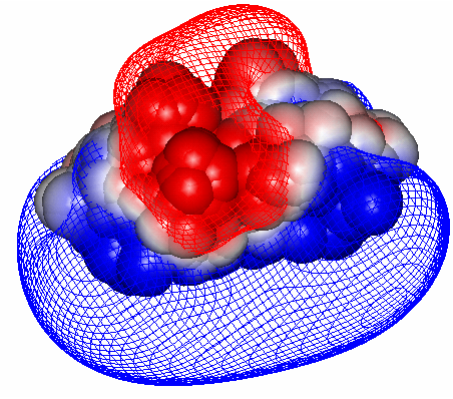
(A) 0.5 M



(B) 0.1 M



(C) 0.02 M



(D) 0.0045 M

Figure 10. 0.5(blue) and -0.5(red) kT/e electrostatic potential contours around BLG AA dimer at pH 5 at ionic strengths (A) 0.5 M, (B) 0.1M, (C) 0.02 M, (D)0.0045M

VI.B.1.3..*Kinetic Analysis*

The strong increase in the intial rate with decrease in ionic strength, and the linear dependence of intial rate on $1/I$ as seen in figure 9 can be explained by considering protein aggregation as a diffusion controlled process. If protein A is diffusing towards any other protein, protein B as depicted in Figure 11, the diffusion trajectory of A is given as:

$$A = \nu_0 \, . \, t \qquad (2)$$

where ν_0 is the velocity and t is time. The probability that the diffusion trajectory of A

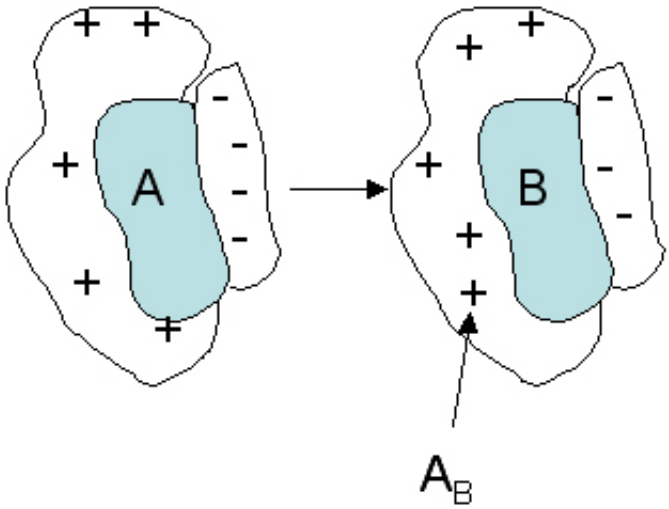


Figure 11. A simplified picture of protein A diffusing towards protein B.

intersects a cross-sectional area of B should be given by A times the cross sectional area density of any other protein. The cross-sectional area density is obtained by multiplying the number density of the proteins n/V in cm^{-3} (where n is the number of proteins and V is the total volume of the system) by A_B (in cm^2). This probability is therefore given by:

$$P = A\frac{n.A_B}{V} \qquad (3)$$

The total probability will be a dimensionless quantity equal to 1, since the probability of effective collisions should be equal to unity.

$$P = A\ \frac{n.A_B}{V} = 1 = v_0 \cdot t \cdot \frac{n.A_B}{V} \qquad (4)$$

The rate of the process is proportional to *1/t*, and solving this equality to get the expression for the rate given by *1/t* reveals that the rate is directly proportional to the cross sectional area A_B assuming υ_0 and volume V are constant.

$$rate \propto \frac{1}{t} \propto A_B \propto R^2 \qquad (5)$$

At low ionic strengths, where the Debye length (K^{-1}) is large compared to the protein radius, the cross sectional area is therefore proportional to K^{-1}, which is directly proportional to *1/I* giving a linear relationship between the initial rate of aggregation and *1/I.* The uncertain geometry of the positive domain in Figure 10D just gives a dependence of $(\kappa^{-1})^a$; in the case of a simple sphere a=2. At high ionic strengths, where, charges are correctly screened and the collisions in the correct orientation is necessary, and a more detailed treatment would be required.

VI.B.1.4.*Sample Variability*

The marked differences in the aggregation rates of the different lots seen in figure 5 could not be ascribed to any of the known characteristics of the proteins. It did not correlate with sample age (ranging from a few months to more than a year), nor with the reported purity of the samples (97% or higher for all lots). We considered the influence of the genetic variants A and B, but the A:B ratio for all lots was close to 1:1. Nevertheless, the resultant observations on the aggregation of the genetic variants, discussed below, provide some insight into the aggregation seen for Sigma "AB" BLG shown in the preceding figure 6.

VI.B.1.5.*Genetic Variance*

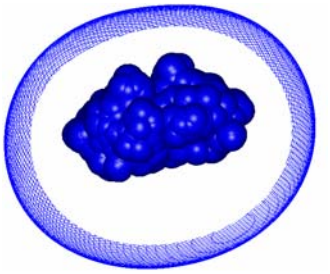
Figure 6 shows turbidimetric response of BLG variants obtained by stopped-flow spectrophotometry. In experiments I and IV 0.5 g/L of BLG A or BLG B solutions respectively, at pH 9.0 were rapidly brought to a target pH of 4.8. The resultant solution at pH 4.8 would have dimers of BLG A or BLG B respectively, as the initial reactants. Results in figure 6 show rapid aggregation for BLG A,(I) and negligible aggregation for BLG B (IV). In II BLG A and B were mixed at pH 9.0 to yield a solution of 1 g/L in total protein concentration and then brought rapidly to pH 4.8. The resultant solution would be expected to have some mixture of dimers of AA, BB and AB as the initial reactants . Significant aggregation was noticed in (II), which was however lower than for (I). In III, BLG A and B were first dissolved separately at pH 6.0 to yield a solution of 1g/L and then mixed at that pH, and then brought to 4.8, resulting in dimers of BLG A and dimers of B only as the initial reactants. Results in figure 6 show negligible aggregation for III.

The observations from figure 6 provide some insight into the aggregation seen for sigma"AB" BLG shown in figure 3 and 9. In (I) of figure 6, the strong aggregation of the AA dimer is consistent with the results of Townsend and Timasheff[84] who reported the presence of AA oligomers substantially higher than dimers. On the other hand, no aggregation is noticed with the BB dimers presumably formed in (IV). The two units of the BB dimer both lack the Asp at position 64, present in BLG A. As will be shown below, these two acidic amino acids lead to significantly larger negative domains in the AA dimer at pH 4.8, which can interact with the large positive domain present at this pH

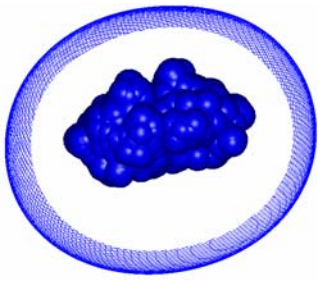
for all possible dimers. The BB dimer's lack of A negative domain at this pH can suppress its self-aggregation. In the case of III where we start with AA and BB dimers as the initial reactants, BB is non-reactive and seems to suppress the aggregation of AA when it is present in suitable quantity. In the case II where AA, AB and BB dimers can exist, the observed aggregation could be due to that of AA and AB. The decrease in aggregation compared to that of (I) could be attributed to several reasons. The presence of BB dimer could be responsible for the suppression of the aggregation of AA and AB. Moreover, AB dimer lacks an aspartatecompared to AA, which might result in reduced aggregation.

The initial rates of aggregation reported in figure 3 and 9 coupled with electrostatic modeling of the genetic variants seem to support the occurrence of some of the aggregation processes noted above. The AA dimer (figure 12 F) with its large positive domain and small but significant negative domain is A unit which can interact with three other AA dimers BY accommodating two negative domains in its large positive domain, and one positive domain for its small negative domain (figure 13); Thus, AA behaves as an X(Y)2 unit (X and Y corresponding to negative and positive domains respectively). Figure 12 (F) also seems to explain the low rate of self-aggregation of BB dimers. This could be due to presence of only an incipient negative domain and a dominating positive domain at pH 4.8. The potential contours for BB at pH 4.8 show a negative domain so small as to be negligible so that this dimer behaves as a Y2 unit (with positive domains only). This not only makes it non-reactive with respect to self-aggregation, but also leads to its ability to suppress the aggregation of AA. Figure 14 depicts the suppression of aggregation of AA dimers by BB dimer, wherein the BB dimer

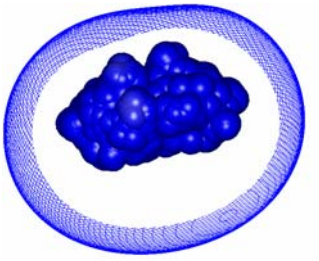
with its significant positive potential suppresses the aggregation of AA dimers and at the same time prevents its own self propagation.



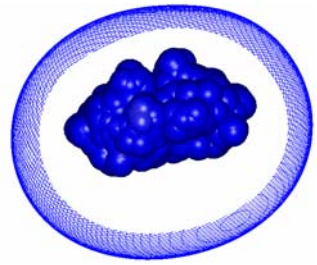
(A) pH3.84 BLGAA



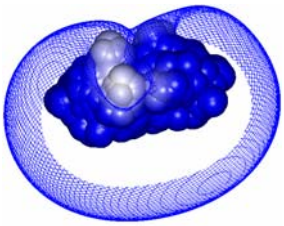
pH3.84 BLGBB



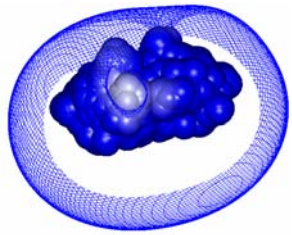
(B) pH4.03 BLGAA



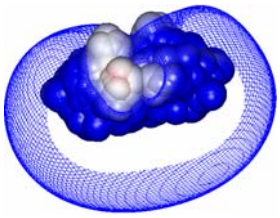
pH4.03 BLGBB



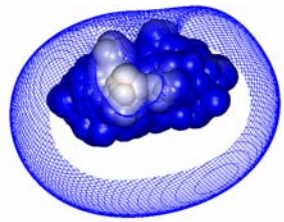
(C) pH4.35 BLGAA



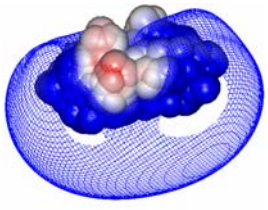
pH4.35 BLGBB

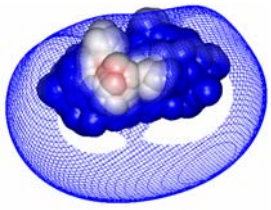


(D) pH4.45 BLGAA

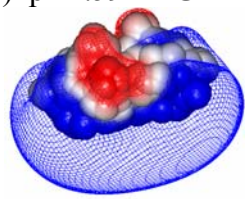


pH4.45 BLGBB

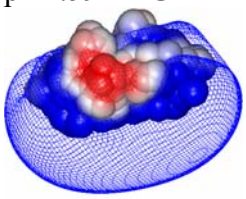




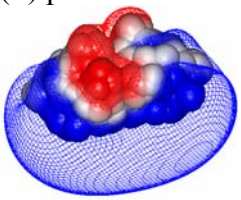
(E) pH4.59 BLGAA



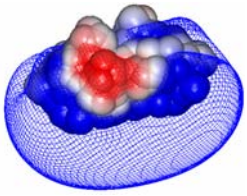
pH4.59 BLGBB



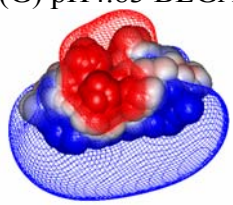
(F) pH4.82 BLGAA



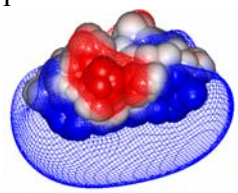
pH4.82 BLGBB



(G) pH4.83 BLGAA

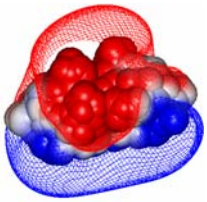


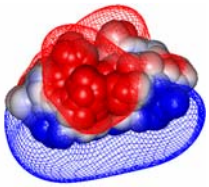
pH4.83 BLGBB



(H) pH4.98 BLGAA

pH4.98 BLGBB





(I) pH5.22 BLGAA

pH5.22 BLGBB

Figure 12: 0.5 (blue) and -0.5(red) kT/e electrostatic potential contours around BLG A and B dimers at ionic strength 0.0045M.

VI.B.1.6..*Aggregation: A Two-Step Process*

Figure 13(A) shows the semi-logarithmic plot of BLG aggregation at pH 4.59 and I = 0.0045M supports the idea of a two step process. A single, first-order process would exhibit a linear fit with a slope proportional to the rate constant. However, the obvious deviation at t = 5min indicates the presence of an initial mechanism with a higher rate constant. The initial rate is extremely sensitive to small changes in pH and depends

strongly on the asymmetric electrostatic domains around the dimer; in particular the magnitude of the negative domain. The second process involves higher-order aggregation of the particles formed in the initial step[80]. The second step does not seem to be sensitive to the electrostatic domains around the dimer, but is related to the global charge, because it is inhibited by the accumulation of the global charge. In fact, the crossing of turbidimetric curves in Figure 13(B) and the crossing of turbidimetric curves (data not shown) from stopped flow spectrophotometer at different pH support the idea of primary and secondary steps having different pH dependencies. This is further supported by the QELS results shown in Figure 7. QELS results show the presence of well defined aggregating species at sizes ranging from 6nm and 200 - 400nm. The 6nm particle size corresponds to the dimer which is the initial reactant and its reactivity determines the initial slope of the turbidity, which forms the step one of the two step process. The higher order particles ranging in the size of 100-200nm could be corresponding to the later process.

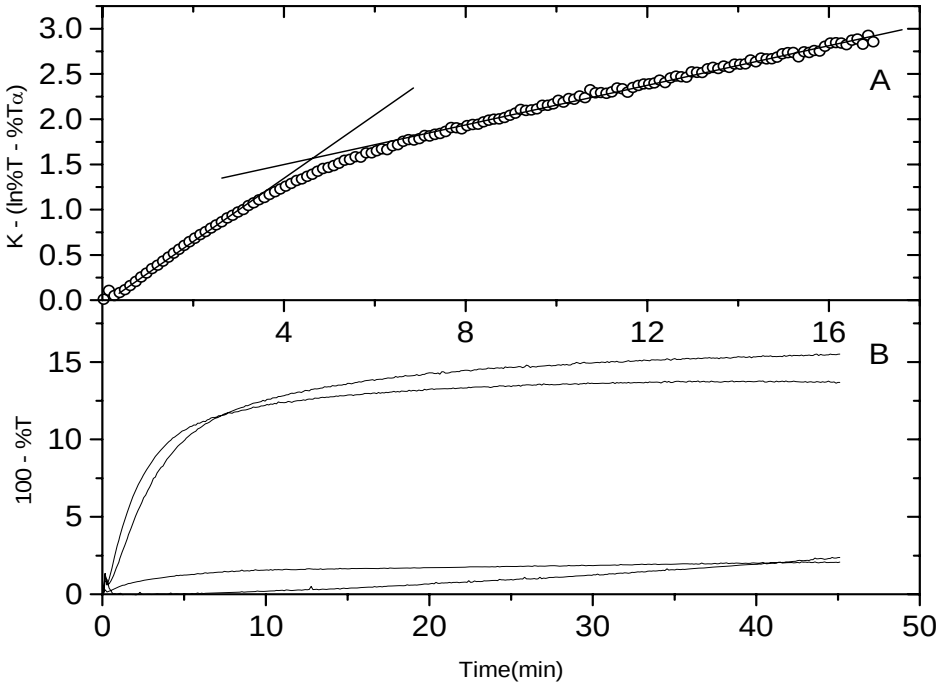


Figure 13(A): Exponential fitting for aggregation data of BLG 1g/L at pH 4.59, and at I= 4.5 mM. Tα is the maximum transmittance value, k is a constant which is equal to ln(%T - %Tα) at time zero. Figure (B): Turbidity vs. time for BLG 1g/L, I = 4.5 mM at pH values from top to bottom: 4.82, 4.59, 5.22 and 3.84

VI.C. INSULIN

VI.C.1.RESULTS AND DISCUSSION

VI.C.1.1.*Effect of pH*

As seen in Figures 14 and 15, compared with Figures 1 and 3, the dependence of insulin aggregation on pH resembles the results obtained for BLG, but with some prominent distinguishing features. As is the case for BLG, the initial increase in turbidity

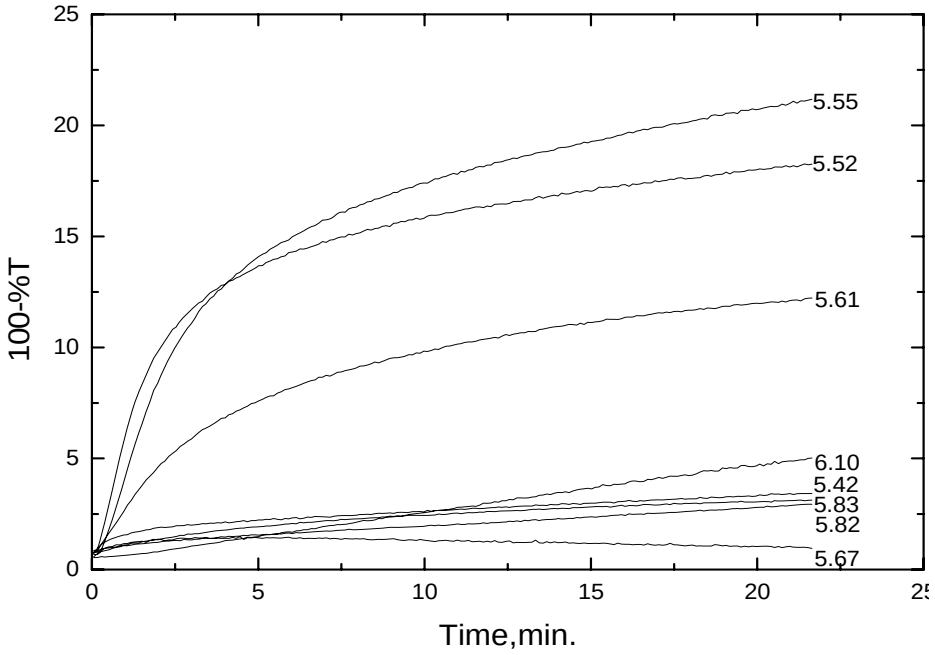


Figure 14. Turbidity vs. time for insulin 0.13g/L, I = 10mM (Sodium acetate) at pH values shown.

shows a marked decrease when the initial pH deviates by +/- 0.1 pH unit from the pH of maximum aggregation, (pH_{max}) ca. 5.5. However, there are a number of significant differences. (1) While $(d\tau/dt)_0$ for BLG becomes zero at pH = pH_{max} + 0.7, zero

aggregation is never seen for insulin in the pH range studied. (2) Comparison of Figure 1 and 14 shows the absence of a plateau in the case of insulin. (3) Comparison of Figure 2 and 15 shows that the pH dependence is slightly symmetric for insulin with a maximum effect of pH below pH_{max} but, in case of BLG, more asymmetric with a maximum effect of pH on the higher side pH_{max}. (4) The maximum in $(d\tau/dt)_0$ for insulin occurs at pH = pI, while it is seen at pH = pI - 0.5 for BLG. It is necessary to consider these marked

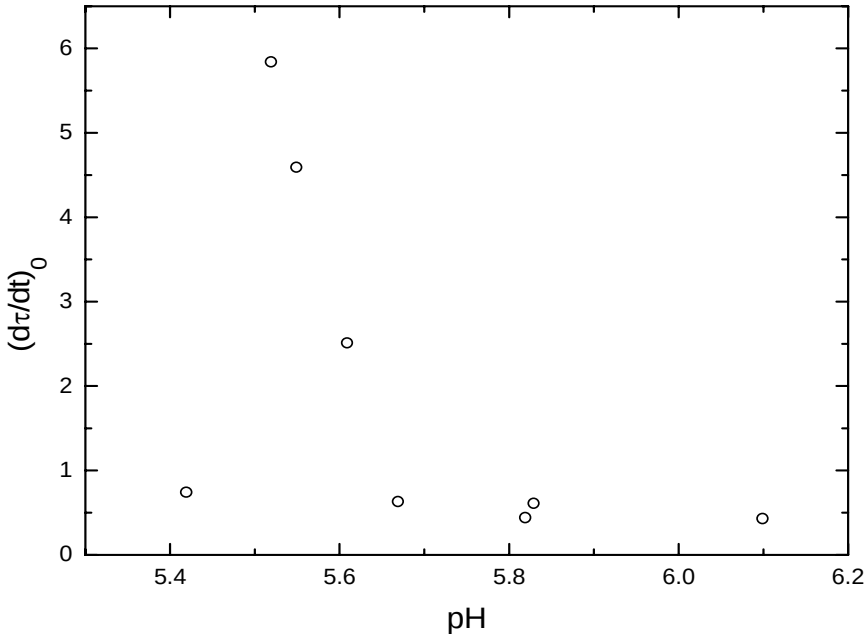


Figure 15. Initial Rate vs. pH for 0.13g/L insulin at 10mM ionic strength.

differences in terms of the nonuniformity of the electrostatic domains for both proteins, and their implications for both multimer formation and aggregation.

The effect of pH on the rate of aggregation of insulin can be distinguished from that of BLG in an alternative way. Figure 16 shows turbidimetric titrations of BLG and insulin, involving the addition of acid to a solution at pH 9.0 ("forward"), or the addition

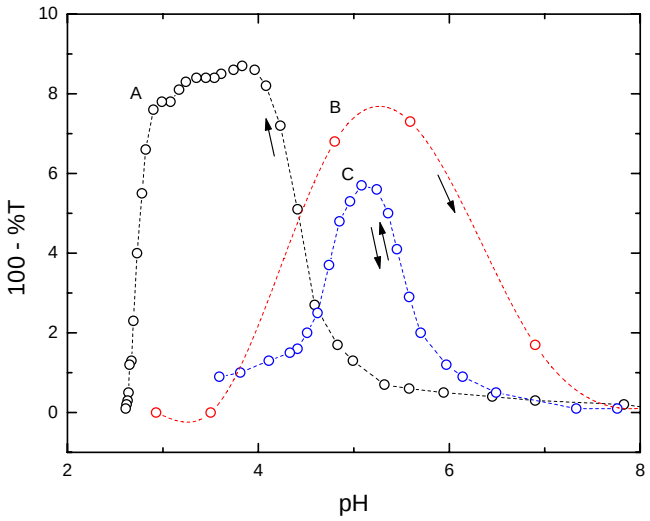


Figure 16. Type 1 titrations of BLG "forward" (starting from pH 9.0) (A); and "reverse" (starting from pH 3.0) (B); and insulin "forward" (C), all at 0.5g/L I = 4.5mM and 10mM NaCl . "Reverse" titration for insulin (not shown) is identical to "forward".

of base to a solution at pH 3.0 ("reverse") (both of these referred to for historical reasons as "Type 1 titrations"). The base titration of insulin (not shown) coincides exactly with its forward titration, i.e. full reversibility. On the other hand the titrations of BLG reveal strong hysterisis, and incomplete reversibility. For example, BLG is strongly aggregated at pH 3.3 in the forward titration, but aggregation does not appear at this pH in the reverse titration, presumably showing the slowness of disaggregation in the forward titration. It might be surprising that maximum turbidity in the forward titration is seen at pH 4 – 3, which is well below $pH_{max.}$; this is because turbidity stops increasing (attains a maxium) when aggregation stops –[$(d\tau/dt)_0$ = 0] which according to Figure 2 is at pH

3.8, and this is exactly where turbidity is a maximum in the "forward titration" of Figure 16. The decrease in turbidity at pH<4 in the forward titrtation corresponds to disaggregation, first slow, then rapid at pH<3. In these titrations, attainment of a turbidity maximum indicates the end of the aggregation process and simultaneous initiation of the disruption of the aggregates due to increase in the repulsive forces. The point of maximum slope at pH ~ 4.5 in Figure 16A is the point of maximum $(d\tau/dt)_0$. In fact, a tangent drawn to any point on the aggregating (right) side of the curve would give the rate of the aggregation and its pH-dependence would be Figure 2 , while a tangent drawn to any point on the other side of the curve would give the rate of disaggregation. Figure 17 portrays how these tangents change with pH. The process of forward aggregation stops at pH 4.0, however aggregation from low pH (Figure 16 B) starts at pH 3.5. For a reversible reaction, both these pH values would have been the same.. Similarly, the derivative of the disaggregating phase of forward titration would give the (pH-dependent) rate of disaggregation. Higher order aggregates formed in forward titration persist to pH 3.0, while a corresponding pH in Figure 2 shows aggregation rates close to zero.

For insulin, forward and backward curves are identical, and both are nearly symmetrical, unlike BLG. The reversibility of insulin aggregation could be due to aggregation in equilibrium with disaggregation[36], i.e. reactants in equilibrium with the products. This pH dependent aggregation behavior as seen both in Figures 2 and 16A (BLG), and Figures 15 and 16B (insulin), arises from the distinctly different electrostatic properties and aggregation mechanisms of insulin and BLG.

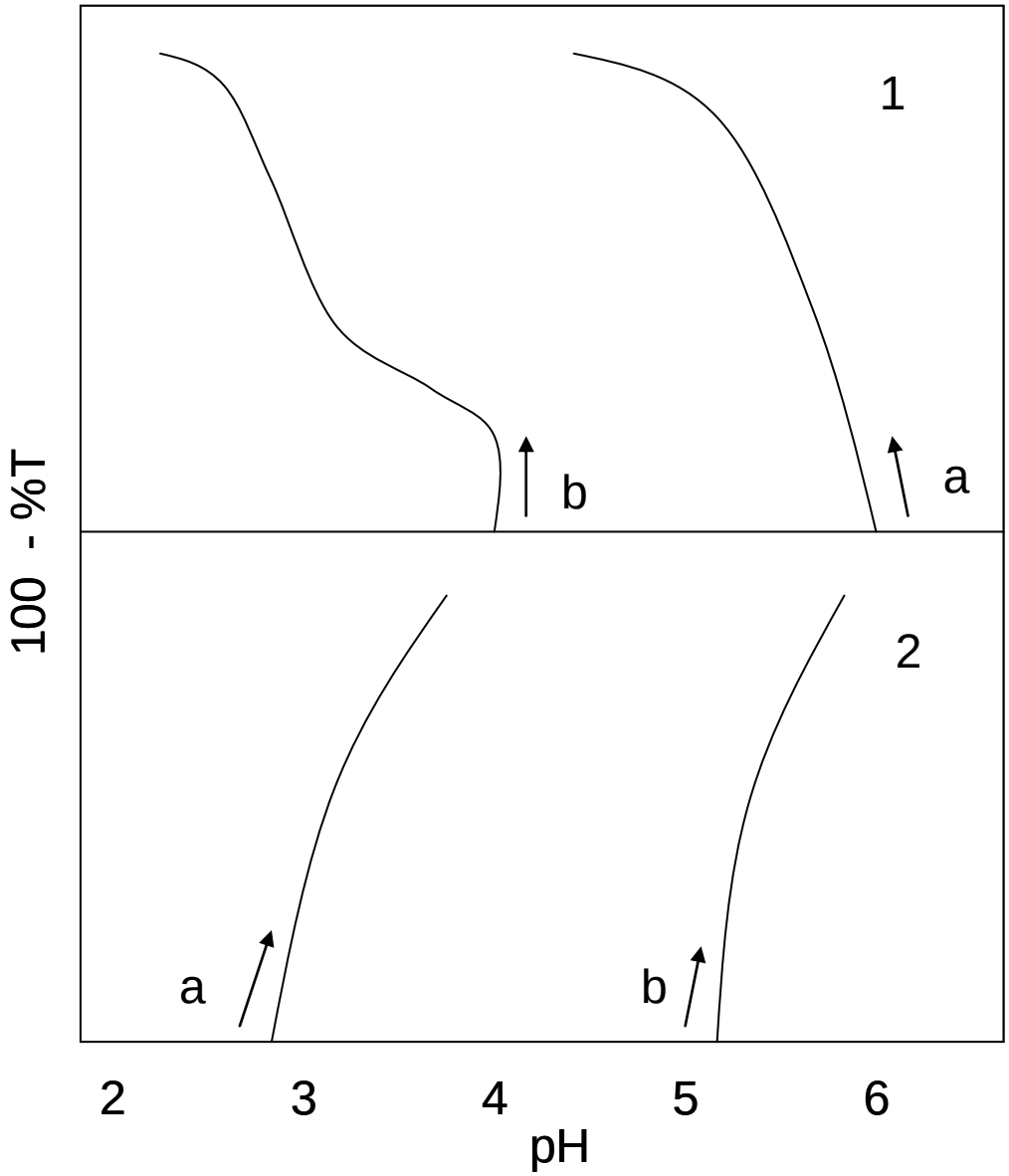


Figure 17. Derivatives of Type 1 titrations of BLG : 1(a) forward aggregation, 1(b) forward disaggregation, 2 (a) reverse disaggregation 2(b) reverse aggregation.

We consider the dimer to be the reactant for aggregation because (1) only dimer and hexamer have been reported in our pH range, and (2) the hexamer has a completely positive core[85]. The electrostatic models of the dimer in Figure 18 show approximately equal positive and negative lobes. These models do not show marked changes in the

electrostatic potential contours in the regime of aggregation, and implicate linear aggregation, with the possibility of a negative domain of one dimer seating itself in the positive pocket of another, and so on. This can be visualized as a linear polymerizing chain of dimers in an (A-B) fashion, all of its subunits linked identically. The process of disaggregation would just involve removing a single oligomer from the aggregate. On the contrary, electrostatic models of BLG in the aggregating regime have a huge positive potential surface coupled with a small negative domain, which facilitates $A\text{-}(B)_2$ type of aggregation. The process of disaggregation for such a three dimensional aggregate would involve breaking many associations some of which are in the interior of an aggregate particle, consequently making the process of disaggregation less reversible.

The electrostatic domains influence multimer formation as well as aggregation. Electrostatic models of insulin reveal complementary positive and negative domains with a strongly preferred orientation for each connection zone, i.e. an orientational energy minimum. On the other hand, asymmetric potential domains seen in electrostatic models of BLG lead to complexes in which there are have multiple orientations of similar energy; this leads to formation of a loose non-specific aggregate. A single preferred binding orientation leads to formation of multimers, while formation of unstructured complexes occurs as a result of numerous binding configurations. Insulin forms well defined reversible, multimers involving monomer, dimer and, hexamer which are in equilibrium. For example, insulin hexamer has a well defined surface with its positive potential patches oriented to its core, and negative surface to the periphery. In contrast, BLG association is open ended, and involves formation of unstructured multimers.

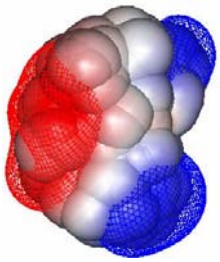
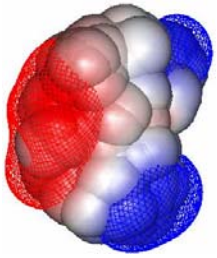
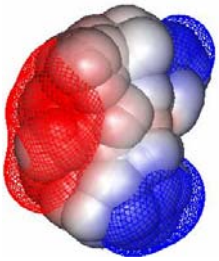
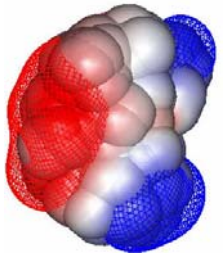
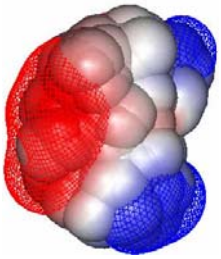
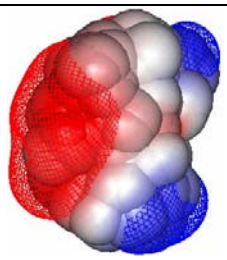
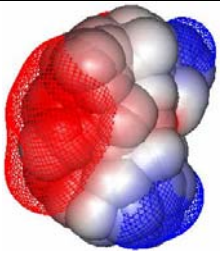
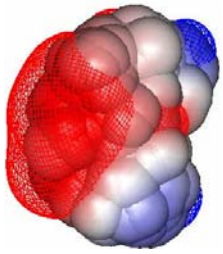
 pH 5.42	 pH 5.52
 pH 5.55	 pH 5.61
 pH 5.67	 pH 5.82
 pH 5.83	 pH 6.10

Figure 18: 0.5(blue) and -0.5(red) kT/e electrostatic potential contours around insulin dimer at ionic strength 0.01M at pH 5.42, 5.52, 5.55, 5.61, 5.67, 5.82, 5.83, and 6.10.

The differences in the aforementioned pH dependent aggregation behavior for insulin vs. BLG can be accounted for as follows. (1) In contrast to insulin, BLG aggregation usually attains a plateau with time. BLG aggregation is highest when the dimer has a net positive charge, and accumulated positive charge prevents indefinite aggregation. For insulin, on the other hand, pH_{max} = pI, where the net global charge would be close to zero, so aggregation occurs without charge accumulation. (2) pH dependence of aggregation in more asymmetric for BLG than for insulin. Asymmetry is seen in BLG when the rate depends very strongly on pH, especially at pH>pHmax. The asymmetry occurs as a result of large changes in the rate of aggregation with small changes in pH.because of the very strong influence of the small negative domain. In contrast, insulin requires symmetric positive and negative domains to aggregate. (3) The maximum rate of initial aggregation for BLG, is at pH_{max}< pI where it has asymmetric electrostatic domains. For insulin pH_{max}= pI, where it has net charge of zero and symmetric electrostatic domains leading to a different aggregation behavior.

VI.C.1.2.*Effect of ionic strength*

The ionic strength dependence of aggregation of insulin studied by Type 1 titrations shows maxium turbidity at 10 mM NaCl. (Figure 19). BLG was not examined in this way because of its irreversibility and complex time-dependence (Figure 5), but it would be expected to show a much lower aggregation at 50 mM and much larger aggregation at 1 mM. In contrast, maximum turbidity was observed in Figure 18 at 10mM NaCl. At ionic strengths below and above 10mM NaCl, turbidity decreased

suggesting a decrease in aggregation rate. The maxima in turbidity shifted to the lower pH's with both increase and decrease in ionic strength. If the symmetry of these type 1 titration curves indicates reversibility, with aggregation increasing as pH is lowered towards pHmax, and disaggregation increasing as pH continues to fall below pHmax, then the maxima in turbidity represents a point where the net aggregation is now zero, because the rates of aggregation and disaggregation are equal. The point of maximum turbidity represents the initiation of disaggregation in a forward titration, and the initiation of aggregation in a reverse titration.

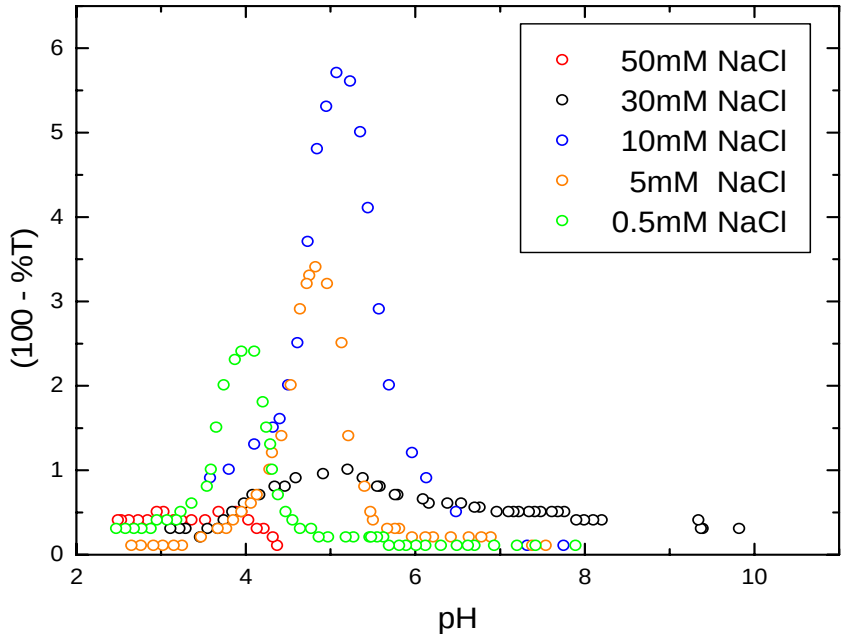


Figure 19. Ionic strength dependence of type I titration for insulin 0.1g/L at I = 0.5mM, 5mM, 10mM, 30mM and, 50mM.

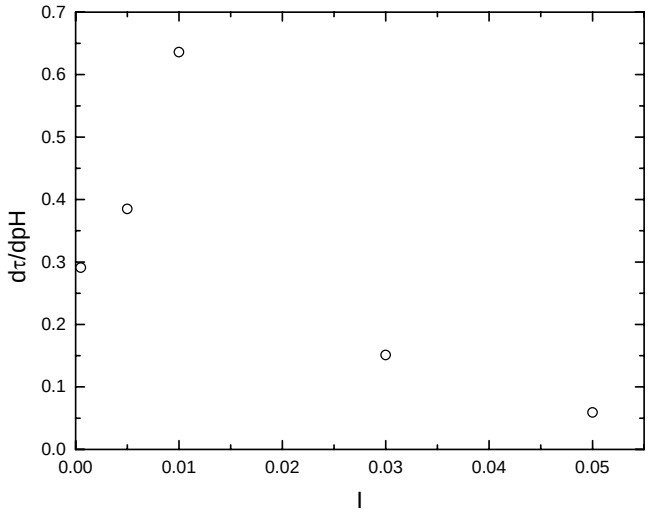
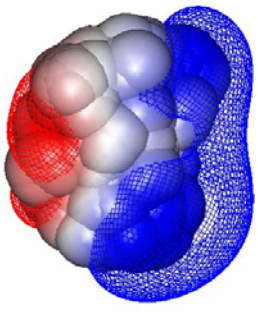
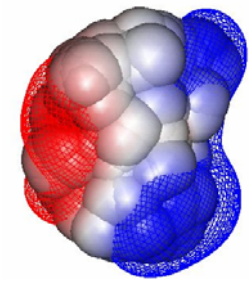


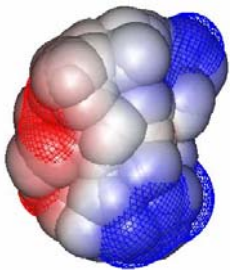
Figure 20. Ionic strength dependence of (dτ/dpH) at the pH of maximum aggregation rate (maximum negative slope in Figure 19) vs. *I*



(A)



(B)



(C)

Figure 21. 0.5(blue) and -0.5(red) kT/e electrostatic potential contours around insulin dimer at (A) pH 5.0, *I* = 0.5mM, (B) pH 5.2, *I* = 10mM and (C) pH 5.0, *I* = 50mM NaCl.

Figure 20 shows dτ/dpH obtained from type 1 titrations of insulin (Figure 19). It is based on the assumption that dτ/dpH ≅ dτ/dt (this would be true if these increases in turbidity would occur even if the pH were held constant at the conditions of each data point. Support for this assumption comes from the fact that the pH at which dτ/dpH is maximum for BLG in the aggregation side of Figure 16A is nearly equal to pH 4.6 where dτ/dt)0 is maximum in Figure 3). Figure 20 is thought to resemble the result for the true ionic strength dependence of the initial rate of aggregation for insulin. dτ/dpH for insulin shows a maximum at pH 5.5, I = 10mM. Maximum aggregation of insulin in 10mM sodium chloride can be considered in terms of the Debye-Huckel screening length κ^{-1}. The distance between the centers of mass of two adjacent dimers *a* and *b* is 3 nm, so a screening length of 3 nm allows interaction between the positive domain of *a* (*a*+) and the negative domain of *b* (*b*-). This corresponds to I = 10 mM. At low salt, e.g. 0.5 mM, κ^{-1} = about 13 nm, so that repulsions, *a*+ interacting with *b*+ and *a*- with *b*- also occur. At high salt, e.g. 50 mM NaCl, κ^{-1} = 1.3 nm, which diminishes the attractive interaction between *a*+ and *b*-. These points are shown in Figure 21, where there is a marked positive potential domain for 5 mM, and completely screened potentials at 50mM But at 10mM, the intermolecular distance in the aggregating species would be equal to the screening length. Allowing the positive potential surface of a dimer to only see the negative surface of another dimer. This interaction could lead to the propagation of a linear chain.

VII. CONCLUSION

The pH dependence of BLG aggregation is markedly asymmetric. Maximum aggregation rates were observed in the pH range 4.3 to 4.8, below isoelectric point of 5.2, so that where positive potential domains are larger than negative ones. This is explained by the hypothesis that such asymmetry makes it possible for each aggregating subunit (BLG dimer) to interact with three other such dimers, the optimal condition for such interaction providing a relatively small but not negligible negative domain. The rate of aggregation strongly increased with decrease in ionic strength I and was found to be nearly linear with $1/I$; suppression of aggregation at high ionic strengths is due to the screening of attractive forces. The aggregation mechanism of BLG can be explained by a two-step process in which the initial reactant (dimer), rapidly forms intermediate clusters (200 nm) and these clusters further associate, rather slowly, to form much larger aggregates (800 nm). The growth of the larger clusters may be limited by the accumulation of excess charge.

The aggregation of insulin is different in many was from that of BLG. (1) Insulin aggregation was maximum at its pI (5.5) where positive and negative domains are of equal magnitude. (2) The ionic strength dependence of insulin aggregation doesn't seem to have a linear relationship with $1/I$, but instead decreases with both increase and decrease in ionic strength. The highest rate of aggregation at I = 10mM can be explained by the ideal Debye-Huckel screening length of the insulin dimer. (3) Type I titrations of insulin are completely reversible, unlike that of BLG which took a hysterisis. This reversibility of insulin can be explained by the symmetric domains of the dimer which

have strong preferential orientation. In case of the BLG which has multiple orientations of similar energy aiding in the formation of loose specific aggregates results in irreversibility. (4) Electrostatic modeling has been used to understand pH and ionic strength dependence of insulin and BLG aggregation and the resulting inter-macromolecular effects related to protein charge heterogeneity.

VIII. FUTURE WORK

- Size exclusion chromatography of BLG and insulin needs to be carried out to determine the size of the predominant aggregating species.

- Aggregation experiments should be carried out with insulin to determine the ionic strength dependence of initial rate of aggregation.

- Light scattering studies of insulin should be done to determine the species involved in insulin aggregation.

- Aggregation studies can be extended to other globular proteins like serum albumin and lysozyme to determine whether electrostatics can be implicated in their aggregation behavior as well.

IX. REFERENCES

1. Spengler, S. J. *Science.* 2000, 287: P. 1221.
2. Odynets, K. S.; Kovalskyy, D. B.; Tokovenko, B. T.; Kornelyuk, A. I. *Biopolimeri i Klitna.* 2004, 20: p. 78.
3. Nakamura, H. A.; Wada, A. *J. Phys. Soc. Jpn.* 1985, 54: p. 4047.
4. Warshel, A.; Levit, M. *J. Mol. Biol.* 1976, 103: p. 227.
5. Schulz, G. E.; Schirmer, R. H. Springer-Verlag: New York, 1979.
6. Carter, P. A.; Wells, J. A. *Nature.* 1988, 332: p. 564.
7. Parsons, S. M.; Raftery, M. A. *Biochemistry II.* 1972, p. 1623.
8. Schutz, C. N.; Warshel, A. *Proteins.* 2001**,** 44: p. 400.
9. Honig, B.; Nicholls, A. *Science.* 1995, 268: p. 1144.
10. Warshel, A.; Aqvist, J. *Annu. Rev. Biophys. Chem.* 1991, 20: p. 267.
11. Bashford, D. *Curr. Opin. Struct. Biol.* 1991, 1: p. 175.
12. Gilson, M. K.; Honig, B. *Proteins.* 1988, 4: p. 7.
13. Wade, R. C. G., R. R.; Ludemann, S. K.; Lounnas, V. *Proc. Natl. Acad. Sci.* 1998, 95: p. 5942.
14. Laberge, M., *Biochim. Biophys, Acta.* 1998, 1386: p. 305.
15. daSilva, F. L. B.; Jonsson, B.; Penfold, R. *Protein Sci.* 2001, 10: p. 1415.
16. Zhang, L. Y.; Gallicchio, e.; Friesner, R. A.; Levy, R. M. *J. comp. Chem.* 2001, 22: p. 591.
17. Simonson, T. *Curr. Opin. Struct. Biol.* 2001, 11: p. 243.

18. Nielsen, J. E.; Andersen, K. V.; Honig, B.; Hooft, R. W. W.; Klebe, G.; Vriend, G. and Wade, R. C. *Protein Eng.* 1999, 12: p. 657.
19. Havranek, J. J.; Harbury, P. B. *Proc. Natl. Acad. Sci.* 1999, 96: p.11145.
20. Kirkwood, J. G. *Biophys. J.* 1934, 80: p. 69.
21. Linderstrom-Lang, K., *Compt. Rend. Trav. Lab.* 1924, 15: p.1.
22. Nozaki, Y.; Tanford, C. *Meth. Enzymol.* 1967, 11: p. 715.
23. Rocchia, W. A., E.; Honig, B. *J. phys. Chem. B.* 2001, 105: p. 6507.
24. Honig, B.; Nicholls, A. *Science.* 1995, 268: p. 1144.
25. Ullmann, G. M. H.; M.; Jensen, A. and Knapp, E. *Proteins.* 2000, 38: p. 301.
26. Elcock, A. H. *J. Mol. Biol. 2001,* 312: p. 885.
27. Cellmer, Troy; Bratko, Dusan; Prausnitz, John M.; Blanch, Harvey. *Biotech and Bioeng,* 2004, 89: p. 78.
28. Goldgaber, Dimitry Y.; Schwarzman, Alexander L.; Eisenberg-Grunberg, Moises. Method of preventing aggregation of amyloid b-protein. PCT Int. Appl. 1995, 53 pp.
29. DuBay Kateri F.; Pawar Amol P.; Chiti Fabrizio.; Zurdo Jesus.; Dobson Christopher M.; Vendruscolo Michele. *J. Mol Biol.* 2004, 341: p 1317.
30. Stoppini, Monica.; Andreola, Alessia.; Foresti, Giovanni.; Bellotti, Vittorio. *Pharmacological Research* 2004, 50: p. 419.
31. Bratko, D.; Blanch, H. W. *J. Chem. Phy.* 2001, 114: p. 561.
32. Cellmer, Troy.; Bratko, Dusan.; Prausnitz, John M.; Blanch, Harvey. *Biotech and Bioeng.* 2004, 89: p. 78.

33. Cellmer, T.; Bratko, D.; Prausnitz, J.M.; Blanch, H. *Biotechnol. bioeng*. 2005, 89: p. 78.

34. Berg, O.G.; Hippel, P. H. *Annu. Rev. Biophys. Biophys. Chem.* 1985, 14: p. 131.

35. Schreiber, G.; Fersht, A. R. *Nature Struct. Biol.* 1996, 3: p. 427.

36. Selzer, T., Albeck, S., Schreiber, G., *Nature Struct Biol.* 2000, 7: p. 537.

37. Elcock, A. H.; Gabdoulline, R.R.; Wade, R.C.; McCammon, J.A. *J. Mol. Biol.* 1999, 291: p. 149.

38. Getzoff, E. D.; Cabelli, D.E.; Fisher, C.L.; Parge, H.E.; Viezzoli, M.E., Banci, L., Hallewell, R.L.*Nature.* 1992, 358: p. 347.

39. Sharp, K.; Fine, R.; Honig, B. *Science.* 1987, 236: p. 1460.

40. Vagedes, P.; Saenger, W.; Knapp, E. W. *Biophy J.* 2002, 83: p. 3066.

41. Vijaykumer et. al., *J. Mol. Biol*. 1998, 278: p. 1015.

42. Schreiber, G.; Buckle, A. M.; Fersht, A. R. *Nat. Struct. Biol*. 1994, 2: p. 945.

43. Wase.; Gabdoulline, A. *J. Mol. Recognit.* 1999, 12: p. 226.

44. Sawyer, Lindsay; Kontopidis, George; Wu, Su-Ying. *Int. J. Food Sci. Tech.* 1999, 34: p. 409.

45. Perez, M. D., and Calvo, M. *J. Dairy Sci.* 1995, 78: p. 978.

46. Macleod, A.; Fedio, W.M.; Ozimek, L. *Milchwissenschaft.* 1995, 50 : p. 666.

47. Kumosinski, T.F.; Timasheff, S.N. *J. Am. Chem. Soc*. 1966. 88: p. 2635.

48. McKenzie, H.A.; Sawyer, W. H. W. H., *Nature* 1967. 214: p. 1101.

49. Sakurai, K.; Oobatake, M.; Goto, Y. *Protein Sci*. 2001, 10: p. 2325.

50. Timasheff, S.N.; Townend, Robert. J. *J. Am. Chem. Soc*. 1961, 83: p. 464.

51. Baldini, Giancarlo.; Beretta, Sabrina.; Chirico, Giuseppe.; Franz, Hermann.; Maccioni, Elisabetta.; Mariani, Paolo.; Spinozzi, Francesco. *Macromolecules.* 1999, 32: p. 6128.
52. Iametti, Stefania.; Scaglioni, Leonardo.; Mazzini, Stefania.; Vecchio, Giuseppe.; Bonomi, Francesco. *J. Agric. Food Chem*. 1998, 46: p. 2159.
53. Verheul, M.; Pedersen, J.S.; Roefss, S.P.; de Kruif, K. G. *Biopolymers* 1999, 49: p. 11.
54. Townsend, Robert.; Timasheff, Serge. N. *J. Am. Chem. Soc*. 1960. 82: p. 3161.
55. McKenzie, H.A.; Sawyer, W. H.; Smith, M.B. *Biochimica et biophysica acta.*. 1967, 147: p. 73.
56. Piazza, R.; Iacopini, S. *Eur. Phy. J*. 2002, **7**: p. 45.
57. Gottschalk, M.; Nilsson., H., *Protein Sci.* 2003. 12: p. 2404.
58. Renard, D.; Lefebvre, J.; Griffin, M.C.A.; Griffin, W. G. *Macromol*. 1998, 22: p. 41.
59. Amyard, P.; Durand, D.; Nicolai. T. *Int. J. Biol.* Macromol. 1996, 22: p. 41.
60. Piazza, R.; Iacopini, S. *Eur. Phy. J*. 2002, **7**: p. 45.
61. Neves-Petersen, M.T.P.; Steffen. B. *Int. J. Biol. Macromol*. 2003, 9: p. 315.
62. Brown, H.; Sanger, F.; Kitai, R. *Biochem. J*. 1955, 60: p. 556.
63. Ryle, AP.; Sanger, F.; Smith, LF.; Kitai, R. *Biochem. J.* 1955, 60: p. 541.
64. Adams, MJ.; Blundell, TL.; Dodson, EJ et al. *Nature.* 1969, 224: p. 491.
65. Steiner, DF.; Oyer, PG. *Proc. Nat. Acad. Sci*. 1967, 57: p. 473.

66. Blundell, TL.; Dodson, GG.; Hodgkin, DC.; Mercola, D. *Adv. Prot. Chem.* 1972, 26: p. 279.

67. Pekar, AH.; Frank, BH. *Biochemistry*. 1972, 11: p. 4013.

68. Goldman, J.; Carpenter, FH. *Biochemistry*. 1974, 13:p. 4566.

69. Jeffrey, PD.; Milthorpe, BK.; Nichol, LW. *Biochemistry*. 1976, 15: p. 4660.

70. Holladay, LA.; Ascoli, M.; Puett, D. *Biochim, Biophys. Acta*. 1977, 494: p. 245.

71. Mark, AE.; Nichol, LW.; Jeffrey, PD. *Biophys. Chem*. 1987, 27: p. 103.

72. Hvidt, S. *Biophys. Chem.* 1991, 39: p. 205.

73. Kadima, W.; Ogendal, L.; Bauer, R.; Kaarsholm, N.; Brodersen, K.; Hansen, JF.; Porting, P. *Biopolymers.* 1993, 371: p. 1165.

74. Strazza, S.; Hunter, R.; Walker, E.; Darnall, DW. *Arch. Biochem. Biophys.* 1985, 238: p. 30.

75. Svboda, I., Brandeburg, D.; Barth, T., Gattner, H-G.; Jiracek, J., Velek, J.; Blaha, I.; Ubik, K.; Kasicka, V., Pospicek, J.; Hrbas, P. *Biol. Chem.* Hoppe-Seyler., 1994, 375: p. 373.

76. Kline, AD.; Justice, RM jr. *Biochemistry*, 1990, 29: p. 2906.

77. Tanford, C.; Kirkwood, J.G. *J. Am. Chem. Soc,* 1957, 79: p. 5333.

78. Nozaki, Y.; Bunville, LG.; Tanford, C. *J. Am. Chem. Soc*, 1959, 81: p. 5523.

79. Tanford, C.; Epstein, J.; *J. Am. Chem. Soc*, 1954, 76: p. 2163.

80. Reddy, Ganta. 2004. Electrostatically driven aggregation of Beta-Lactoglobulin and effects of added polyelectrolytes. M.S. Thesis.

81. Seyrek, E.; Dubin, PL.; Tribet, C.; Gamble, EA. *Biomacromolecules*, 2003, 4: p. 273.

82. Mathews, Van. Holde.; Ahern, K. *Biochemistry*. Third Edition ed. 2000.

83. Flory, P.J., *Principles of Polymer Chemistry*. 1953, 672.

84. Townsend, Robert.; Timasheff, Serge. N. *J. Am. Chem. Soc*, 1960, 82: p. 3168.

85. Diepen, M. 1996. Crystallographic studies of modified insulin. PhD. Thesis.